On Stein's method and perturbations

Barbour, A D ; Cekanavicius, V ; Xia, A

Abstract: Stein's (1972) method is a very general tool for assessing the quality of approximation of the distribution of a random element by another, often simpler, distribution. In applications of Stein's method, one needs to establish a Stein identity for the approximating distribution, solve the Stein equation and estimate the behaviour of the solutions in terms of the metrics under study. For some Stein equations, solutions with good properties are known; for others, this is not the case. Barbour and Xia (1999) introduced a perturbation method for Poisson approximation, in which Stein identities for a large class of compound Poisson and translated Poisson distributions are viewed as perturbations of a Poisson distribution. In this paper, it is shown that the method can be extended to very general settings, including perturbations of normal, Poisson, compound Poisson, binomial and Poisson process approximations in terms of various metrics such as the Kolmogorov, Wasserstein and total variation metrics. Examples are provided to illustrate how the general perturbation method can be applied.

Originally published at:
Barbour, A D; Cekanavicius, V; Xia, A (2007). On Stein's method and perturbations. ALEA Latin American Journal of Probability and Mathematical Statistics, 3:31-53.

On Stein's method and perturbations

A. D. Barbour[*], Vydas Čekanavičius[†] and Aihua Xia[‡]
Universität Zürich, Vilnius University
and University of Melbourne

Abstract

Stein's (1972) method is a very general tool for assessing the quality of approximation of the distribution of a random element by another, often simpler, distribution. In applications of Stein's method, one needs to establish a Stein identity for the approximating distribution, solve the Stein equation and estimate the behaviour of the solutions in terms of the metrics under study. For some Stein equations, solutions with good properties are known; for others, this is not the case. Barbour & Xia (1999) introduced a perturbation method for Poisson approximation, in which Stein identities for a large class of compound Poisson and translated Poisson distributions are viewed as perturbations of a Poisson distribution. In this paper, it is shown that the method can be extended to very general settings, including perturbations of normal, Poisson, compound Poisson, binomial and Poisson process approximations in terms of various metrics such as the Kolmogorov, Wasserstein and total variation metrics. Examples are provided to illustrate how the general perturbation method can be applied.

Keywords: perturbation method, normal distribution, jump diffusion process, Poisson distribution, compound Poisson distribution, Poisson process, point

[*]Angewandte Mathematik, Winterthurerstrasse 190, CH–8057 ZÜRICH, Switzerland: `a.d.barbour@math.unizh.ch`

[†]Department of Mathematics and Informatics, Vilnius University, Naugarduko 24, Vilnius 03225, Lithuania: `vydas.cekanavicius@mif.vu.lt`

[‡]Department of Mathematics and Statistics, University of Melbourne, Vic 3010, Australia: `xia@ms.unimelb.edu.au`

process, total variation norm, Kolmogorov distance, Wasserstein distance, local distance.

1 Introduction

Many applications of Stein's (1972) method, when approximating the distribution $\mathcal{L}(W)$ of a random element W of a metric space $\mathcal{X}$ by a probability distribution π, are accomplished broadly as follows. The aim is to estimate $\mathbb{E}h(W) - \pi(h)$ for each member h of a family of test functions $\mathcal{H}$, where $\pi(h) := \int h d\pi$. To do this, one finds a normed space $\mathcal{G}$ and an appropriate Stein operator $\mathcal{A}$ on $\mathcal{G}$ characterizing π; $\mathcal{A}\colon \mathcal{G} \to \mathcal{F} \subset \mathbb{R}^{\mathcal{X}}$, for some $\mathcal{F} \supset \mathcal{H}$, must be such that $\pi(\mathcal{A}g) = 0$ for all g in $\mathcal{G}$, and that π is the unique probability distribution for which this is the case. 'Appropriate' in this context means that an inequality of the form

$$|\mathbb{E}\{(\mathcal{A}g)(W)\}| \;\le\; \varepsilon \|g\|_{\mathcal{G}}, \qquad g \in \mathcal{G}, \tag{1.1}$$

can be established, for some (small) ε. Finally, for each $h \in \mathcal{H}$, find a function $g_h \in \mathcal{G}$ satisfying the Stein equation

$$\mathcal{A}g_h \;=\; h - \pi(h). \tag{1.2}$$

Then it follows from (1.1) that

$$|\mathbb{E}h(W) - \pi(h)| \;\le\; \varepsilon \|g_h\|_{\mathcal{G}}. \tag{1.3}$$

Hence, if it can be shown that

$$\|g_h\|_{\mathcal{G}} \;\le\; C\|h\|_{\mathcal{F}}, \tag{1.4}$$

for some norm $\|\cdot\|_{\mathcal{F}}$ on $\mathcal{F}$, we can conclude that

$$d_{\mathcal{H}}(\mathcal{L}(W), \pi) \;\le\; C\varepsilon \sup_{h\in\mathcal{H}} \|h\|_{\mathcal{F}}, \tag{1.5}$$

where, for any two distributions P and Q on $\mathcal{X}$,

$$d_{\mathcal{H}}(P, Q) \;:=\; \sup_{h\in\mathcal{H}} |P(h) - Q(h)|. \tag{1.6}$$

Thus, if (1.2) and (1.4) are satisfied, it is enough for the $d_{\mathcal{H}}$-approximation of $\mathcal{L}(W)$ by π to establish the inequality (1.1); in this sense, Stein's method for π can be said to work for the distance $d_{\mathcal{H}}$. Distances of this form include the total variation distance d_{TV}, with $\mathcal{H}$ the set of functions bounded by 1, and the Wasserstein distance d_W, with $\mathcal{H}$ the Lipschitz functions with slope bounded by 1.

Probabilistic inequalities of the form (1.1) can be derived by a variety of techniques, including Stein's exchangeable pair approach, the generator method and Taylor expansion. However, the analytic inequality (1.4) can prove to be a stumbling block, especially if a reasonably small value of C is desired, unless π happens to be a particularly convenient distribution. For $\mathcal{X} = \mathbb{R}$, the normal and Poisson distributions lead to simple versions of (1.4). However, when introducing Stein's method for compound Poisson distributions, Barbour, Chen & Loh (1992) were only able to prove analogous inequalities with satisfactory values of C for distributions for which the generator method was applicable, and this represents a strong restriction on the compound Poisson family. The class of amenable compound Poisson distributions was subsequently extended in Barbour & Xia (1999), where a perturbation technique was introduced, which enabled the good properties of the solutions of the Poisson operator to be carried over to those of the Stein equations for neighbouring compound Poisson distributions. Their approach was taken further in Barbour & Čekanavičius (2002) and in Čekanavičius (2004). Here, we show that the perturbation idea can be applied not just in the Poisson setting, but in great generality. One consequence is that the range of compound Poisson distributions whose solutions have good properties can be further extended, but the scope of possible applications is much wider. In particular, there is no need to restrict attention to random variables on the real line; distributions and random elements on quite general spaces can be considered.

The perturbation method is discussed in the general terms in Section 2. Theorem 2.1 shows how to find the solution g_h in (1.2) for $\mathcal{A} = \mathcal{A}_1$, when $\mathcal{A}_1$ is close enough to a 'nice' Stein operator $\mathcal{A}_0$, and the probability measure π_0 associated with $\mathcal{A}_0$ has $\mathrm{supp}\,(\pi_0) = \mathcal{X}$; the theorem also gives the inequality corresponding to (1.4). Theorem 2.4 gives conditions under which Stein's method works, but which do not assume the support condition, and Theorem 2.5 allows a further slight relaxation, which is particularly relevant to approximation of random variables using the Kolmogorov distance. In Section 3, a number of specific examples are given, some of which are illustrated from the point of view of application in Section 4.

As indicated above, there are various ways in which an inequality (1.1) relevant in any particular setting may be derived. This means that the choice of operator $\mathcal{A}_1$, and of the corresponding approximating probability measure π_1, is frequently dictated by the problem under consideration in a more or less natural way. The choice of $\mathcal{A}_0$ is more a matter of chance. If $\mathcal{A}_1$ is not itself one of the operators for which the solutions to (1.2) are known to satisfy an inequality of the form (1.4), then one looks for an $\mathcal{A}_0$ which is, and which is not too far away from $\mathcal{A}_1$. Such an operator need not exist. In order for our perturbation approach to be successful, it is necessary for the contraction inequality (2.8) to be satisfied, and this limits the set of operators which can be considered as perturbations of any given $\mathcal{A}_0$, for the purposes of our theorems.

2 Formal approach

Let $\mathcal{X}$ be a Polish space, and $\mathcal{G}$ a linear subspace of the functions $g\colon \mathcal{X} \to \mathbb{R}$ equipped with a norm $\|\cdot\|_{\mathcal{G}}$. Suppose that π_0 is a probability measure on $\mathcal{X}$ with $\operatorname{supp}(\pi_0) = \mathcal{X}_0 \subset \mathcal{X}$. Define

$$\begin{aligned}
\mathcal{F} &:= \{f\colon \mathcal{X} \to \mathbb{R},\ \pi_0(|f|) < \infty\};\\
\mathcal{F}_0 &:= \{f \in \mathcal{F}\colon f(x) = 0 \text{ for all } x \notin \mathcal{X}_0\};\\
\mathcal{F}' &:= \{f \in \mathcal{F}\colon \pi_0(f) = 0\}; \qquad \mathcal{F}'_0 \ := \ \mathcal{F}_0 \cap \mathcal{F}',
\end{aligned}$$

and let P_0 be the projection from $\mathcal{F}$ onto $\mathcal{F}'_0$ given by

$$P_0 f \ := \ f\mathbf{1}_{\mathcal{X}_0} - \pi_0(f)\mathbf{1}_{\mathcal{X}_0},$$

where, here and subsequently, $\mathbf{1}_A$ denotes the indicator function of the set A, and multiplication of functions is to be understood pointwise. Now let $\|\cdot\|$ be a norm on $\mathcal{F}$, set

$$\overline{\mathcal{F}} \ := \ \{f \in \mathcal{F}\colon \|f\| < \infty\},$$

and define $\overline{\mathcal{F}}_0 := \overline{\mathcal{F}} \cap \mathcal{F}_0$, $\overline{\mathcal{F}}' := \overline{\mathcal{F}} \cap \mathcal{F}'$, $\overline{\mathcal{F}}'_0 := \overline{\mathcal{F}} \cap \mathcal{F}'_0$; we shall require that $\|\cdot\|$ is such that

$$P_0\colon \overline{\mathcal{F}} \to \overline{\mathcal{F}}'_0. \tag{2.1}$$

We also assume that $\overline{\mathcal{F}}$ is a determining class of functions for probability measures on $\mathcal{X}$ (Billingsley 1968, p. 15).

We now suppose that there is a 'nice' Stein operator $\mathcal{A}_0$ characterizing π_0. By this, we mean that

$$\mathcal{A}_0 \colon \mathcal{G} \to \mathcal{F}_0', \tag{2.2}$$

and also that it is possible to define a right inverse

$$\mathcal{A}_0^{-1} \colon \overline{\mathcal{F}}_0' \;\to\; \mathcal{G}_0 \;:=\; \{g \in \mathcal{G} \colon g(x) = 0 \text{ for all } x \notin \mathcal{X}_0\},$$

satisfying

$$\mathcal{A}_0(\mathcal{A}_0^{-1} f) \;=\; f \quad \text{for all} \quad f \in \overline{\mathcal{F}}_0'; \tag{2.3}$$

$$\|\mathcal{A}_0^{-1} P_0 f\|_{\mathcal{G}} \;\le\; A\|f\|, \qquad f \in \overline{\mathcal{F}}, \tag{2.4}$$

for some $A < \infty$. Note that (2.2) means that

$$\pi_0(\mathcal{A}_0 g) = 0 \quad \text{for all} \quad g \in \mathcal{G}. \tag{2.5}$$

On the other hand, in view of (2.3), if π is any probability measure on $\mathcal{X}_0$ such that $\pi(\mathcal{A}_0 g) = 0$ for all $g \in \mathcal{G}$, then $\pi(f) = 0$ for all $f \in \overline{\mathcal{F}}_0'$, meaning that $\pi(f) = \pi_0(f)$ for all $f \in \overline{\mathcal{F}}_0$, and hence for all $f \in \overline{\mathcal{F}}$. Since $\overline{\mathcal{F}}$ is a determining class, $\pi = \pi_0$, and $\mathcal{A}_0$ characterizes π_0 through (2.5).

In the setting of the introduction, for $h \in \mathcal{H} \subset \mathcal{F}_0$ a family of test functions, we have $h(x) - \pi_0(h) = (P_0 h)(x)$ for $x \in \mathcal{X}_0$, so that we can take $g_h = \mathcal{A}_0^{-1} P_0 h$ and obtain (1.2), in view of (2.3). Inequality (2.4) is just (1.4) for $\mathcal{A}_0$, with f in place of h. Hence, because of (1.5), Stein's method for π_0 based on (1.1) (with $\mathcal{A}_0$ in place of $\mathcal{A}$) works for distances based on families $\mathcal{H}$ of test functions whose norms are uniformly bounded. Our interest here is in extending this to probability measures π_1 characterized by generators $\mathcal{A}_1$ which are close to $\mathcal{A}_0$.

So let π_1 be a finite signed measure on $\mathcal{X}$ with $\pi_1(\mathcal{X}) = 1$, and such that $|\pi_1|(|f|) < \infty$ for all $f \in \overline{\mathcal{F}}$. Let $\mathcal{A}_1$ be a Stein operator for π_1, meaning that $\mathcal{A}_1 \colon \mathcal{G} \to \mathcal{F}_1'$, where

$$\mathcal{F}_1' \;:=\; \{f \colon \mathcal{X} \to \mathbb{R};\; |\pi_1|(|f|) < \infty,\; \pi_1(f) = 0\},$$

so that

$$\pi_1(\mathcal{A}_1 g) \;=\; 0 \quad \text{for all} \quad g \in \mathcal{G}; \tag{2.6}$$

set $\mathcal{U} = \mathcal{A}_1 - \mathcal{A}_0$, and assume also that

$$\mathcal{U}\mathcal{A}_0^{-1}P_0 \colon \overline{\mathcal{F}} \to \overline{\mathcal{F}}. \tag{2.7}$$

The key assumption which ensures that $\mathcal{A}_1$ can fruitfully be thought of as a perturbation of $\mathcal{A}_0$ is that

$$\|\mathcal{U}\mathcal{A}_0^{-1}P_0\| \ =: \ \gamma \ < \ 1. \tag{2.8}$$

Remark. Having to satisfy the condition (2.8) significantly limits the choice of distributions π_1 whose Stein equations can be treated as perturbations of that for π_0. This is clearly illustrated in the examples of the next section.

Theorem 2.1 *With the above definitions, suppose that assumptions (2.1)–(2.4) and (2.6)–(2.8) are satisfied. Then the operator*

$$\mathcal{B} \ := \ \mathcal{A}_0^{-1}P_0 \sum_{j\geq 0}(-1)^j(\mathcal{U}\mathcal{A}_0^{-1}P_0)^j \colon \ \overline{\mathcal{F}} \ \to \ \mathcal{G}_0 \tag{2.9}$$

is well defined, and

$$\|\mathcal{B}\| \ \leq \ A/(1-\gamma); \qquad \|\mathcal{U}\mathcal{B}\| \ \leq \ \gamma/(1-\gamma). \tag{2.10}$$

Furthermore, for $f \in \overline{\mathcal{F}}$ and for all $x \in \mathcal{X}_0$,

$$(\mathcal{A}_1\mathcal{B}f)(x) - (P_1 f)(x) \ = \ c(f) \ = \ \pi_1(f) - \pi_0(f) + \pi_0(\mathcal{U}\mathcal{B}f), \tag{2.11}$$

where $P_1 f = f - \pi_1(f)\mathbf{1}$; here, $\mathbf{1} = \mathbf{1}_{\mathcal{X}}$. In particular, if $\mathcal{X}_0 = \mathcal{X}$, we have $c(f) = 0$, so that $\mathcal{B}$ is a right inverse of $\mathcal{A}_1$ on $\mathcal{F}_1' \cap \overline{\mathcal{F}}$.

Proof. The first part is immediate from (2.4) and (2.8), from the properties of $\mathcal{A}_0^{-1}$ and from (2.7). It is then also immediate that

$$(\mathcal{A}_0 + P_0\mathcal{U})\mathcal{B}f \ = \ P_0 f, \quad f \in \overline{\mathcal{F}}.$$

Hence, for $f \in \overline{\mathcal{F}}$, we have

$$\begin{aligned} \mathcal{A}_1\mathcal{B}f &= (\mathcal{A}_0 + P_0\mathcal{U} + (I - P_0)\mathcal{U})\mathcal{B}f \\ &= P_0 f + (\mathcal{U}\mathcal{B}f)\mathbf{1}_{\mathcal{X}_0^c} + \pi_0(\mathcal{U}\mathcal{B}f)\mathbf{1}_{\mathcal{X}_0}, \end{aligned} \tag{2.12}$$

so that, for $x \in \mathcal{X}_0$,

$$(\mathcal{A}_1\mathcal{B}f)(x) - (P_1f)(x) \;=\; \pi_1(f) - \pi_0(f) + \pi_0(\mathcal{U}\mathcal{B}f) \;=:\; c(f). \qquad (2.13)$$

For the constant $c(f)$, note that, from (2.6) with $\mathcal{B}f$ for g and from (2.12), we have

$$\begin{aligned} 0 \;&=\; \pi_1(f\mathbf{1}_{\mathcal{X}_0}) - \pi_1(\mathcal{X}_0)\pi_0(f) + \pi_1((\mathcal{U}\mathcal{B}f)\mathbf{1}_{\mathcal{X}_0^c}) + \pi_0(\mathcal{U}\mathcal{B}f)\pi_1(\mathcal{X}_0) \\ &=\; \pi_1(f) - \pi_1(f\mathbf{1}_{\mathcal{X}_0^c}) - \pi_0(f) + \pi_1(\mathcal{X}_0^c)\pi_0(f) + \pi_1((\mathcal{U}\mathcal{B}f)\mathbf{1}_{\mathcal{X}_0^c}) \\ &\qquad\qquad + \pi_0(\mathcal{U}\mathcal{B}f)(1 - \pi_1(\mathcal{X}_0^c)). \end{aligned}$$

This implies, from the first part of the theorem, that

$$c(f) \;=\; \pi_1(f) - \pi_0(f) + \pi_0(\mathcal{U}\mathcal{B}f) \;=\; 0$$

if $\mathcal{X}_0^c = \emptyset$, and

$$c(f) \;=\; \pi_1(f\mathbf{1}_{\mathcal{X}_0^c}) - \pi_1(\mathcal{X}_0^c)\pi_0(f) - \pi_1((\mathcal{U}\mathcal{B}f)\mathbf{1}_{\mathcal{X}_0^c}) + \pi_0(\mathcal{U}\mathcal{B}f)\pi_1(\mathcal{X}_0^c) \quad (2.14)$$

otherwise. □

Remark. If $\mathcal{X}_0 = \mathcal{X}$, then it follows from Theorem 2.1 that

$$\mathcal{A}_1\mathcal{B}h \;=\; P_1h \;=\; h - \pi_1(h)$$

for test functions $h \in \mathcal{H} \subset \overline{\mathcal{F}}$. Hence, for such h, the function $g_h := \mathcal{B}h$ satisfies (1.2), where $\mathcal{A}$ is replaced by $\mathcal{A}_1$ and π by π_1. It then follows from (2.10) that $\|g_h\|_{\mathcal{G}} \le A(1-\gamma)^{-1}\|h\|$, so that (1.4) is satisfied with $C = A/(1-\gamma)$, and hence Stein's method for π_1 based on (1.1) (with $\mathcal{A}_1$ for $\mathcal{A}$) works for distances $d_{\mathcal{H}}$ derived from bounded families of test functions.

If $\mathcal{X}_0 \neq \mathcal{X}$, the inequalities (2.10) are still satisfied, so that (1.4) is still true with $C = A/(1-\gamma)$ if $g_h = \mathcal{B}h$. However, this choice of g_h now gives only an approximate solution to (1.2):

$$(\mathcal{A}_1 g_h)(x) \;=\; h(x) - \pi_1(h) + c(h), \qquad x \in \mathcal{X}_0. \qquad (2.15)$$

This is still enough to show that Stein's method works for π_1 based on (1.1) (with $\mathcal{A}_1$ for $\mathcal{A}$), as is demonstrated in Theorem 2.4 below. To make the connection, we first need two more lemmas.

The first concerns the size of $|c(f)|$. This can be controlled in a number of ways, two of which are given in the following lemma. For any finite signed measure π and any $A \subset \mathcal{X}$, we define

$$\kappa(\pi, A) \;:=\; \sup_{\{f \in \overline{\mathcal{F}}:\, \|f\| \le 1\}} |\pi|(\hat{f}\mathbf{1}_A), \tag{2.16}$$

where $\hat{f}(x) := |f(x) - \pi_0(f)|$.

Lemma 2.2 *For $f \in \overline{\mathcal{F}}$, we have*

$$\begin{aligned} &\textit{(i)} \qquad |c(f)| \;\le\; \frac{2|\pi_1|(\mathcal{X}_0^c)}{1-\gamma}\,\|f\|_\infty; \\ &\textit{(ii)} \qquad |c(f)| \;\le\; \frac{\kappa(\pi_1, \mathcal{X}_0^c)}{1-\gamma}\,\|f\|. \end{aligned}$$

Proof. The proof is immediate from (2.14) and (2.16). □

The second lemma translates (2.15) into an inequality bounding the difference $|\pi(f)-\pi_1(f)|$ in terms of $|\pi(\mathcal{A}_1\mathcal{B}f)|$, for a general probability measure π on $\mathcal{X}$.

Lemma 2.3 *Under the conditions of Theorem 2.1, if π is any probability measure on $\mathcal{X}$, then, for any $f \in \overline{\mathcal{F}}$, we have*

$$|\pi(f) - \pi_1(f)| \;\le\; |\pi(\mathcal{A}_1\mathcal{B}f)| + \begin{cases} 2(1-\gamma)^{-1}\{|\pi_1|(\mathcal{X}_0^c) + \pi(\mathcal{X}_0^c)\}\,\|f\|_\infty; \\ (1-\gamma)^{-1}\{\kappa(\pi_1, \mathcal{X}_0^c) + \kappa(\pi, \mathcal{X}_0^c)\}\,\|f\|. \end{cases}$$

Proof. It follows from (2.12) that

$$\begin{aligned} \pi(\mathcal{A}_1\mathcal{B}f) &= \pi(P_0 f) + \pi((\mathcal{U}\mathcal{B}f)\mathbf{1}_{\mathcal{X}_0^c}) + \pi_0(\mathcal{U}\mathcal{B}f)\pi(\mathcal{X}_0) \\ &= \pi(f) - \pi(f\mathbf{1}_{\mathcal{X}_0^c}) - \pi_0(f)(1 - \pi(\mathcal{X}_0^c)) \\ &\qquad + \pi((\mathcal{U}\mathcal{B}f)\mathbf{1}_{\mathcal{X}_0^c}) + \pi_0(\mathcal{U}\mathcal{B}f)(1 - \pi(\mathcal{X}_0^c)) \\ &= \{\pi(f) - \pi_1(f)\} + c(f) - \pi(f\mathbf{1}_{\mathcal{X}_0^c}) \\ &\qquad + (\pi_0(f) - \pi_0(\mathcal{U}\mathcal{B}f))\pi(\mathcal{X}_0^c) + \pi((\mathcal{U}\mathcal{B}f)\mathbf{1}_{\mathcal{X}_0^c}). \end{aligned}$$

Hence, and using (2.16), the lemma follows. □

This lemma gives the information that we need, when deriving distributional approximations in terms of the measure π_1. Let $\mathcal{H} \subset \overline{\mathcal{F}}$ be any collection of test functions which forms a determining class for probability measures on $\mathcal{X}$. Then define the metric $d_{\mathcal{H}}$ on finite signed measures ρ, σ on $\mathcal{X}$, by

$$d_{\mathcal{H}}(\rho,\sigma) \;:=\; \sup_{h\in\mathcal{H}} |\rho(h) - \sigma(h)|. \tag{2.17}$$

In the special case where $\mathcal{H} := \{f \in \overline{\mathcal{F}} \colon \|f\| \le 1\}$, we write $d_{\overline{\mathcal{F}}}$ for $d_{\mathcal{H}}$. The following theorem shows that Stein's method for π_1 based on (2.18) works for the distance $d_{\overline{\mathcal{F}}}$, even when $\mathcal{X}_0 \neq \mathcal{X}$.

Theorem 2.4 *Suppose that the conditions of Theorem 2.1 are satisfied, and write $g_f^0 := \mathcal{A}_0^{-1} P_0 f$ for all $f \in \overline{\mathcal{F}}$. Then, if*

$$|\pi(\mathcal{A}_1 g_f^0)| \;\le\; \varepsilon \|g_f^0\|_{\mathcal{G}} \quad \textit{for all} \quad f \in \overline{\mathcal{F}}, \tag{2.18}$$

it follows that

$$d_{\overline{\mathcal{F}}}(\pi,\pi_1) \;\le\; (1-\gamma)^{-1}\{A\varepsilon + \varepsilon'(\pi,\pi_1)\},$$

where

$$\varepsilon'(\pi,\pi_1) \;:=\; \min\{2(|\pi_1|(\mathcal{X}_0^c) + \pi(\mathcal{X}_0^c))F, \kappa(\pi_1,\mathcal{X}_0^c) + \kappa(\pi,\mathcal{X}_0^c)\},$$

and

$$F := \sup_{\{f\in\overline{\mathcal{F}}\colon \|f\|\le 1\}} \|f\|_\infty. \tag{2.19}$$

Proof. In fact, let $\tilde f = \sum_{j\ge 0} (-1)^j (\mathcal{U}\mathcal{A}_0^{-1} P_0)^j f$. Then (2.18) together with (2.8) and (2.4) imply that

$$|\pi(\mathcal{A}_1 \mathcal{B} f)| = |\pi(\mathcal{A}_1 g_{\tilde f}^0)| \;\le\; \varepsilon \|g_{\tilde f}^0\|_{\mathcal{G}} \;\le\; A\varepsilon \|\tilde f\| \;\le\; \frac{A\varepsilon}{1-\gamma}\|f\|.$$

Thus the conclusion follows immediately from Lemma 2.3 and from the definition of $d_{\overline{\mathcal{F}}}$. □

Note that (2.18) is a weakening of what would normally be required for (1.1), inasmuch as the inequality is only needed for the functions g_f^0, which, being the solutions to the Stein equation for the 'nice' operator $\mathcal{A}_0$, may well be known in advance to have good properties.

Theorem 2.4 is applied most simply when π is the distribution of some random element W, for which it can be shown that

$$|\mathbb{E}(\mathcal{A}_1 g)(W)| \;\le\; \sum_{j=1}^{l} \varepsilon_j c_j(g), \quad g \in \mathcal{G}. \tag{2.20}$$

Here, the quantities ε_j are to be computed using W alone, and the function g enters only through the constants $c_j(g)$. If the norm $\|\cdot\|$ on $\mathcal{F}$ can be chosen in such a way that the $c_j(g_f^0)$ can be bounded by a multiple of $\|f\|$ for any $f \in \overline{\mathcal{F}}$, then Theorem 2.4 can be invoked.

The choice of norms on $\mathcal{F}$ for which this procedure can be carried through depends very much on the structure of the random variable W: see Section 4. Broadly speaking, for the more stringent norms, the contraction condition (2.8) is harder to satisfy; on the other hand, there are then fewer functions having finite norm, and so the inequality (2.18) is easier to establish. Take, for example, standard normal approximation, with $\mathcal{G}$ the space of bounded real functions with bounded first and second derivatives, endowed with the norm

$$\|g\|_{\mathcal{G}} \;:=\; \|g\|_\infty + \|g'\|_\infty + \|g''\|_\infty\,, \tag{2.21}$$

and with $\mathcal{A}_0$ the Stein operator given by

$$(\mathcal{A}_0 g)(x) \;=\; g'(x) - xg(x), \quad g \in \mathcal{G}. \tag{2.22}$$

Here, it is possible, in many central limit settings, to derive an inequality of the form (1.1):

$$|\mathbb{E}(\mathcal{A}_0 g)(W)| \le \varepsilon \|g\|_{\mathcal{G}}$$

for some ε, as, for example, in Chen & Shao (2005, p. 5). Now, for $g_f^0 = \mathcal{A}_0^{-1} P_0 f$ with $\|f'\|_\infty < \infty$, we have $\|(g_f^0)''\|_\infty \le 4\|f'\|_\infty$ by Proposition 5.1 (c)(i) and (iii) with $y = g_f^0$, so that inequality (1.4) is satisfied with

$$\|f\|^{(1)} \;:=\; \|f\|_\infty + \|f'\|_\infty \tag{2.23}$$

as norm on $\mathcal{F}$. This, in turn, leads to corresponding approximations with respect to the distance $d_{\overline{\mathcal{F}}} =: d^{(1)}$, from (1.5).

In the usual central limit context, there is typically no hope of taking the argument further, and choosing $\mathcal{H} = \overline{\mathcal{F}}$ for the supremum norm $\|\cdot\|_\infty$ in place of $\|\cdot\|^{(1)}$ on $\mathcal{F}$. This is not because the perturbation argument would fail, but because there can usually be no inequality of the form $|\mathbb{E}(P_0 f)(W)| \le \varepsilon \|f\|_\infty$ for all $f \in \overline{\mathcal{F}}$, unless ε is rather large; this is because the supremum of the left hand side is then just the *total variation* distance between $\mathcal{L}(W)$ and the standard normal distribution, and this is not necessarily small under the usual conditions for the central limit theorem. More is, however, possible with some extra restrictions: see Cacoullos *et al.* (1994) and Example 4.1.

The distance $d^{(1)}$ is not the one most commonly used for measuring the accuracy of approximation in the central limit theorem. Here, it is usual to work with the Kolmogorov distance d_K, which is of the form defined in (2.17), with the set of test functions

$$\mathcal{H}^K \;:=\; \{\mathbf{1}_{(-\infty,a]} \colon a \in \mathbb{R}\}.$$

For these test functions, it can in many central limit applications be established, albeit with rather more effort, that $|\mathbb{E}(\mathcal{A}_0 g_h)(W)|$ is bounded, uniformly for $h \in \mathcal{H}^K$, by a quantity of the form $k\varepsilon$ for some $k < \infty$ and ε reflecting the closeness of $\mathcal{L}(W)$ and the standard normal distribution. This in turn, with (1.2), implies error estimates for standard normal approximation, measured with respect to Kolmogorov distance.

Now the set $\mathcal{H}^K$ forms a subset of $\overline{\mathcal{F}}$, when the supremum norm is taken on $\mathcal{F}$, and the perturbation arguments leading to Lemma 2.3 can still be applied successfully, for Stein operators $\mathcal{A}_1$ suitably close to $\mathcal{A}_0$. However, in order to deduce distance estimates as in Theorem 2.4, it is necessary to be able to bound $|\mathbb{E}(\mathcal{A}_0 g_f^0)(W)|$ not only for $f \in \mathcal{H}^K$, but also for any f of the form $f := (\mathcal{U}\mathcal{A}_0^{-1}P_0)^j h$, where $h \in \mathcal{H}^K$ and $j \ge 1$, since these functions are used to make up the function $\tilde{f}$ introduced in the proof of Theorem 2.4. Now these functions f are not typically in the set $\mathcal{H}^K$. However, it can at least be shown that both g_h^0 and $(g_h^0)'$ are uniformly bounded for $h \in \mathcal{H}^K$. For some operators $\mathcal{A}_1$, this is enough to be able to conclude that

$$\sup_{h \in \mathcal{H}^K} \|\mathcal{U} g_h\|^{(1)} \;<\; \infty.$$

It is then possible to apply the following result, in which the Stein operator $\mathcal{A}_0$ is now quite general.

Theorem 2.5 *Suppose that the conditions of Theorem 2.1 are satisfied, and that $\mathcal{H}$ is any family of test functions with $H := \sup_{h\in\mathcal{H}} \|h\|_\infty < \infty$, and such that $g_h^0 := \mathcal{A}_0^{-1} P_0 h$ is well defined for $h \in \mathcal{H}$, satisfying $\mathcal{A}_0 g_h^0 = P_0 h$ and $\mathcal{U} g_h^0 \in \overline{\mathcal{F}}$. Assume further that*

$$\gamma_{\mathcal{H}} \ := \ H^{-1} \sup_{h\in\mathcal{H}} \|\mathcal{U} g_h^0\| \ < \ \infty. \tag{2.24}$$

Then, if π is such that

$$\sup_{h\in\mathcal{H}} |\pi(\mathcal{A}_1 g_h^0)| \ \le \ H\varepsilon_1 \tag{2.25}$$

and

$$|\pi(\mathcal{A}_1 g_f^0)| \ \le \ \varepsilon_2 \|g_f^0\|_{\mathcal{G}}, \quad f \in \overline{\mathcal{F}}, \tag{2.26}$$

it follows that

$$d_{\mathcal{H}}(\pi, \pi_1) \ \le \ H \left\{ \varepsilon_1 + \frac{\gamma_{\mathcal{H}} A \varepsilon_2}{1-\gamma} + \frac{\varepsilon(\pi,\pi_1)}{1-\gamma} \right\},$$

where $\varepsilon(\pi,\pi_1) := \kappa(\pi_1, \mathcal{X}_0^c) + \kappa(\pi, \mathcal{X}_0^c)$.

Proof. Once again, much as in the proof of Theorem 2.4, we note that

$$|\pi(\mathcal{A}_1 \mathcal{B} h)| \ \le \ \sum_{j\ge 0} |\pi(\mathcal{A}_1 \mathcal{A}_0^{-1} P_0 (\mathcal{U}\mathcal{A}_0^{-1} P_0)^j\, h)| \ = \ |\pi(\mathcal{A}_1 g_h^0)| + \sum_{j\ge 1} |\pi(\mathcal{A}_1 g_{f_j}^0)|, \tag{2.27}$$

where $f_j := (\mathcal{U}\mathcal{A}_0^{-1} P_0)^j\, h$, $j \ge 1$. Now, for $h \in \mathcal{H}^K$,

$$\|f_1\| \ = \ \|\mathcal{U} g_h^0\| \ \le \ H\gamma_H,$$

by (2.24), and then, by (2.8),

$$\|f_j\| \ \le \ \gamma^{j-1} H \gamma_H, \qquad j \ge 2.$$

Hence, from (2.27), (2.4), (2.25) and (2.26), it follows that

$$\sup_{h\in\mathcal{H}^K} |\pi(\mathcal{A}_1 \mathcal{B} h)| \ \le \ H\varepsilon_1 + \sum_{j\ge 1} \varepsilon_2 A \gamma^{j-1} H \gamma_H,$$

and the theorem now follows from Lemma 2.3. □

In particular, if $\mathcal{A}_0$ is the Stein operator for normal approximation given in (2.22), and taking the norm $\|\cdot\|^{(1)}$, Theorem 2.5 can be applied with $\mathcal{H} = \mathcal{H}^K$; in circumstances in which the conditions (2.24)–(2.26) are satisfied, this leads to estimates of the error in approximating the distribution π of a random variable W by π_1, measured with respect to Kolmogorov distance. In particular, the estimates (2.25) and (2.26) relating to the distribution of W are of a kind which can often be verified in practice; see Section 4.

3 Examples

In the first two examples, the sets $\mathcal{X}_0$ and $\mathcal{X}$ are the same, so that the elements in the bounds involving probabilities of the set $\mathcal{X}_0^c$ make no contribution. The first of these is purely for illustration, since properties of the Stein equation for the perturbed distribution could be obtained directly.

Example 3.1. In this example, we consider approximation by the probability distribution $\pi_1 := t_{m,\psi}$ on $\mathbb{R}$, with density

$$p_{m,\psi}(x) \;=\; k_{m,\psi}(1 + x^2/m)^{-(m+1)\psi/2}\, e^{-(1-\psi)x^2/2}, \quad x \in \mathcal{X} := \mathbb{R},$$

where $k_{m,\psi}$ is an appropriate normalizing constant. This family of densities interpolates between the standard normal ($\psi = 0$) and Student's t_m distribution ($\psi = 1$) distribution, as ψ moves from 0 to 1; m is classically a positive integer. We take for $\mathcal{G}$ the space of bounded real functions with bounded derivatives, endowed with the norm

$$\|g\|_{\mathcal{G}} \;:=\; \|g\|_\infty + \|g'\|_\infty\,.$$

An appropriate Stein operator $\mathcal{A}_1$ for $t_{m,\psi}$ is given by

$$(\mathcal{A}_1 g)(x) \;=\; g'(x) - x\left\{(1-\psi) + \frac{\psi(m+1)}{m+x^2}\right\} g(x), \quad g \in \mathcal{G}; \tag{3.1}$$

this follows because $p_{m,\psi}(x)$ is an integrating factor for the right hand side of (3.1), and hence, for any $g \in \mathcal{G}$,

$$\int_{-\infty}^{\infty} (\mathcal{A}_1 g)(x) p_{m,\psi}(x)\, dx \;=\; [g(x)p_{m,\psi}(x)]_{-\infty}^{\infty} \;=\; 0,$$

so that (2.6) is satisfied. Now, at least for small enough ψ, $\mathcal{A}_1$ could be thought of as a perturbation of the standard normal distribution, with Stein operator

$$(\mathcal{A}_0 g)(x) \;=\; g'(x) - xg(x), \quad g \in \mathcal{G},$$

discussed above, whose properties are well documented: see, for example, Chen & Shao (2005, Lemmas 2.2 and 2.3). Rather than take the standard normal for π_0, we actually prefer to perturb from a normal distribution $\mathcal{N}(0,(1-\psi)^{-1})$. This has Stein operator

$$(\mathcal{A}_0 g)(x) \;=\; g'(x) - (1-\psi)xg(x), \quad g \in \mathcal{G}, \tag{3.2}$$

which gives

$$(\mathcal{U}g)(x) \;=\; -x\,\frac{\psi(m+1)}{m+x^2}\,g(x), \quad g \in \mathcal{G}.$$

The properties of $\mathcal{A}_0^{-1}$ are as given in Proposition 5.1, with y replaced by g. For the supremum norm on $\mathcal{F}$, we find that assumptions (2.1)–(2.4) and (2.6)–(2.7) are satisfied, and that

$$\begin{aligned} \sup_x |x(\mathcal{A}_0^{-1}P_0 f)(x)| &\;\le\; 2(1-\psi)^{-1}\|f\|; \\ \|\mathcal{U}\mathcal{A}_0^{-1}P_0\| &\;\le\; 2\psi(1-\psi)^{-1}(1+\tfrac{1}{m}) \;=:\; \gamma, \end{aligned}$$

from Proposition 5.1 (b)(iii). Condition (2.8) is satisfied if $\gamma < 1$, in which case Theorem 2.4 shows that Stein's method works.

Note, however, that Student's t_m distribution itself is too far from the normal for this perturbation argument to be applied, since then $\psi = 1$, and so $\gamma = \infty$.

For bounded functions f with bounded derivative, it follows from Proposition 5.1 (c)(iv) that

$$\sup_x |x(\mathcal{A}_0^{-1}P_0 f)'(x)| \;\le\; \frac{3}{1-\psi}\,\|f'\|_\infty\,.$$

This translates into a bound for $\|\mathcal{U}\mathcal{A}_0^{-1}P_0 f\|^{(1)}$, and (2.8) is then satisfied for all ψ small enough. As for normal approximation, bounding $\mathbb{E}\{(\mathcal{A}_0 g)(W)\}$ by a linear combination of $\|g\|_\infty$, $\|g'\|_\infty$ and $\|g''\|_\infty$ may be a much more reasonable prospect than using only $\|g\|_\infty$ and $\|g'\|_\infty$, and these quantities are themselves all bounded by multiples of $\|f\|^{(1)}$, for $g = \mathcal{A}_0^{-1}P_0 f$ and

$f \in \overline{\mathcal{F}}^{(1)} := \{f \in \mathcal{F}\colon \|f\|^{(1)} < \infty\}$: see Proposition 5.1 (c)(i)–(iii), with $y = g$. In such cases, $d^{(1)}$-approximation is a consequence.

To deduce Kolmogorov distance using Theorem 2.5, note that, for $\mathcal{H} = \mathcal{H}^K$,

$$\begin{aligned}\gamma_{\mathcal{H}} &= \sup_{h \in \mathcal{H}^K} \|\mathcal{U} g_h^0\|^{(1)} \\ &\le 2\frac{\psi}{1-\psi}\left(1 + \frac{1}{m}\right)\left\{1 + \frac{1}{\sqrt{m}} + \frac{1}{4}\sqrt{2\pi(1-\psi)} + \frac{1}{2}(1-\psi)\sqrt{m}\right\},\end{aligned}$$

from Proposition 5.1 (a)(i)–(iii). If an approximation with respect to $d^{(1)}$ can be obtained from Theorem 2.4, then the estimate used in (2.18) can be used also in (2.26), and the main further obstacle is thus to verify condition (2.25).

Example 3.2. Our second example also concerns a perturbation of the normal distribution, but now to a distribution π_1, whose Stein operator is not so easy to handle directly. This time, we take for $\mathcal{G}$ the space of real functions g with $g(0) = 0$ and having bounded first and second derivatives, endowed with the norm

$$\|g\|_{\mathcal{G}} := \|g'\|_\infty + \|g''\|_\infty.$$

As Stein operator $\mathcal{A}_1$, we fix $\alpha > 0$ and take the expression

$$(\mathcal{A}_1 g)(x) = g''(x) - xg'(x) + \alpha\{g(x+z) - g(x)\}, \quad g \in \mathcal{G}, \tag{3.3}$$

which can be viewed as a perturbation of the Stein operator

$$(\mathcal{A}_0 g)(x) = g''(x) - xg'(x), \quad g \in \mathcal{G},$$

characterizing the standard normal distribution. This operator is equivalent to that given in (2.22), and the properties of $\mathcal{A}_0^{-1}$ are given in Proposition 5.1, with $y = g'$ and $\psi = 0$. The distribution π_1 is that of the equilibrium of a jump–diffusion process X, with unit infinitesimal variance, and having jumps of size z at rate α.

Once again, taking the supremum norm on $\mathcal{F}$, it is easy to check that assumptions (2.1)–(2.4) and (2.6)–(2.7) are satisfied, and since

$$\|(\mathcal{A}_0^{-1} P_0 f)'\|_\infty \le \sqrt{2\pi}\, \|f\|_\infty,$$

from Proposition 5.1 (b)(i), it follows that

$$\|\mathcal{U}\mathcal{A}_0^{-1}P_0\| \;\le\; \sqrt{2\pi}\,z\alpha. \tag{3.4}$$

Hence, from Theorem 2.4, Stein's method works for π_1 if $\gamma = \sqrt{2\pi}z\alpha < 1$; an estimate of the form (2.18) is all that is needed.

As above, the supremum norm may be more difficult to exploit in practice than the norm $\|\cdot\|^{(1)}$. Here, for $f \in \overline{\mathcal{F}}$, and writing $g_f^0 = \mathcal{A}_0^{-1}P_0 f$, we have

$$|(\mathcal{U}g_f^0)'(x)| \;\le\; \alpha\int_0^z |(g_f^0)''(x+t)|\,dt \;\le\; 4\alpha z\|f\|_\infty,$$

from Proposition 5.1 (b)(ii), and Theorem 2.4 can be applied if α is small enough that $\gamma = (4+\sqrt{2\pi})z\alpha < 1$.

For Kolmogorov approximation, note that, for $\mathcal{H} = \mathcal{H}^K$,

$$\gamma_{\mathcal{H}} \;=\; \sup_{h\in\mathcal{H}^K}\|\mathcal{U}g_h^0\|^{(1)} \;\le\; (1+\sqrt{2\pi}/4)\,z\alpha\,,$$

by Proposition 5.1 (a)(i)–(ii). Once again, the main effort in addition to $d^{(1)}$–approximation is to verify (2.25) of Theorem 2.5.

Note that we are also free to perturb from other normal distributions. If we choose to centre at the mean αz of π_1, we can do so by writing

$$(\mathcal{A}_1 g)(x) \;=\; g''(x) - (x-\alpha z)g'(x) + \alpha\{g(x+z) - g(x) - zg'(x)\}, \quad g\in\mathcal{G},$$

with the first two terms the Stein operator for the normal distribution $\mathcal{N}(\alpha z, 1)$. The third, perturbation term can be bounded by $2\alpha z^2\|f\|_\infty$, and its derivative by $\alpha z^2\|f'\|_\infty$ (Proposition 5.1 (b)(ii)–(iii)), enabling (2.8) to be satisfied for $\|f\|^{(1)}$ for a larger range of α, if z is small enough. It is also possible to begin with $\mathcal{N}(\alpha z, 1+\alpha z^2/2)$, correcting for both mean and variance.

It is also possible to generalize the class of perturbed measures by replacing the term $\alpha(g(x+z)-g(x))$ corresponding to Poisson jumps of rate α and magnitude z by a more general Lévy process, taking instead $\int\{g(x+z)-g(x)\}\,\alpha(dz)$, for a suitable measure α.

Example 3.3. As our third example, considered already in Barbour & Xia (1999) and in Barbour & Čekanavičius (2002), we consider (signed) compound Poisson distributions π_1 on $\mathbb{Z}$, the set of all integers, as perturbations of Poisson distributions on $\mathbb{Z}_+ := \{0, 1, 2, \cdots\}$. We begin with π_1 as the compound Poisson distribution $\mathrm{CP}(\lambda, \mu)$ on $\mathbb{Z}_+$, the distribution of $\sum_{l\ge 1} lN_l$, where $N_1, N_2, \ldots$ are independent, and $N_l \sim \mathrm{Po}\,(\lambda\mu_l)$; $m_1 := \sum_{l\ge 1} l\mu_l$ is assumed to be finite. In this case, we have $\mathcal{X} = \mathcal{X}_0 = \mathbb{Z}_+$. With $\mathcal{G}$ the space of bounded functions $g\colon \mathbb{N} \to \mathbb{R}$, endowed with the supremum norm, a suitable Stein operator for π_1 is given by

$$(\mathcal{A}_1 g)(j) \;=\; \lambda \sum_{l\ge 1} l\mu_l g(j+l) - jg(j), \quad j \ge 0, \tag{3.5}$$

considered as a perturbation of the Stein operator

$$(\mathcal{A}_0 g)(j) \;=\; \lambda m_1 g(j+1) - jg(j), \quad j \ge 0; \tag{3.6}$$

this means that

$$(\mathcal{U}g)(j) \;=\; \lambda \sum_{l\ge 1} l\mu_l \{g(j+l) - g(j+1)\}, \quad j \ge 0. \tag{3.7}$$

Taking the supremum norm on $\mathcal{F}$, assumptions (2.1)–(2.4) and (2.6)–(2.7) are satisfied; and since, from the well-known properties of the solution of the Stein Poisson equation,

$$\|\Delta(\mathcal{A}_0^{-1} P_0 f)\|_\infty \;\le\; \frac{2}{\lambda m_1}\, \|f\|_\infty, \tag{3.8}$$

where $\Delta g(j) := g(j+1) - g(j)$, it follows that

$$\|\mathcal{U}\mathcal{A}_0^{-1} P_0\| \;\le\; 2\lambda \sum_{l\ge 1} l(l-1)\mu_l/(\lambda m_1) \;=\; 2m_2/m_1,$$

where $m_2 = \sum_{l\ge 1} l(l-1)\mu_l$. Hence (2.8) is satisfied if $m_2/m_1 < 1/2$, and Theorem 2.4 can then be invoked. Note that, in this setting, it is reasonable to work in terms of the supremum norm, since total variation approximation may genuinely be accurate.

There are nonetheless other distances that are useful. Two such are the Wasserstein distance d_W, defined for measures P and Q on $\mathbb{Z}$ by

$$d_W(P, Q) \;:=\; \sup_{f \in \mathrm{Lip}_1} |P(f) - Q(f)|,$$

where $\mathrm{Lip}_1 := \{f\colon \mathbb{Z} \to \mathbb{R};\ \|\Delta f\|_\infty \le 1\}$, and the point metric d_{pt} defined by

$$d_{\mathrm{pt}}(P,Q) \;:=\; \max_{j\in\mathbb{Z}} |P\{j\} - Q\{j\}|,$$

which has application when proving local limit theorems.

For Wasserstein distance, it is natural to begin with the semi-norm $\|f\| := \|f\|_W := \|\Delta f\|_\infty$ on $\mathcal{F}$, which becomes a norm when restricted to $\overline{\mathcal{F}}'$. The arguments in Section 2 go through in this modified setting very much as before; the only practical differences are that one needs to check that $P_0\mathcal{U}\mathcal{A}_0^{-1}$ maps $\overline{\mathcal{F}}'_0$ into itself, and to replace the condition (2.8) by

$$\gamma \;:=\; \|P_0\mathcal{U}\mathcal{A}_0^{-1}\| \;<\; 1. \tag{3.9}$$

For the Poisson operator $\mathcal{A}_0$ given in (3.6), it is known that

$$\begin{aligned} &\|g_f^0\|_\infty \;\le\; \|P_0 f\|_W \;=\; \|f\|_W; \quad \|\Delta g_f^0\|_\infty \;\le\; 1.15(\lambda m_1)^{-1/2}\|f\|_W;\\ &\|\Delta^2 g_f^0\|_\infty \;\le\; 2(\lambda m_1)^{-1}\|f\|_W, \end{aligned} \tag{3.10}$$

whenever $f \in \overline{\mathcal{F}}$ and $g_f^0 := \mathcal{A}_0^{-1}P_0 f$ [Barbour and Xia (2005)]. Hence, for $\mathcal{A}_1$ as in (3.5) and $f \in \overline{\mathcal{F}}'_0$, it follows from (3.7) that

$$\|P_0\mathcal{U}g_f^0\|_W \;=\; \|\Delta P_0\mathcal{U}g_f^0\|_\infty \;\le\; \lambda\sum_{l\ge1} l(l-1)\mu_l\|\Delta^2 g_f^0\|_\infty \;\le\; 2(m_2/m_1)\|f\|_W,$$

so that $P_0\mathcal{U}\mathcal{A}_0^{-1}$ indeed maps $\overline{\mathcal{F}}'_0$ into itself, and $\gamma = \|P_0\mathcal{U}\mathcal{A}_0^{-1}\| \le 2m_2/m_1$. Thus (3.9) is satisfied for $m_2/m_1 < 1/2$, and the perturbation approach can then be invoked.

For the point metric, we take the l_1–norm $\|f\| := \|f\|_1 := \sum_{j\in\mathbb{Z}} |f(j)|$ on $\mathcal{F}$. For $f \in \overline{\mathcal{F}}_0$ and $g_f^0 = \mathcal{A}_0^{-1}P_0 f$, we have

$$\|g_f^0\|_\infty \;\le\; (\lambda m_1)^{-1}\|f\|_1; \qquad \|\Delta g_f^0\|_1 \;=\; \sum_{j\ge1} |\Delta g_f^0(j)| \;\le\; 2(\lambda m_1)^{-1}\|f\|_1; \tag{3.11}$$

both inequalities are consequences of the proof of the second inequality in Barbour, Holst & Janson (1992, Lemma 1.1.1). Hence, from (3.7), it follows immediately that

$$
\begin{aligned}
\|\mathcal{U}g_f^0\|_1 &= \sum_{j\ge 0} |(\mathcal{U}g_f^0)(j)| \\
&\le \lambda \sum_{l\ge 1} l\mu_l \sum_{j\ge 0} \sum_{s=1}^{l-1} |\Delta g_f^0(j+s)| \\
&\le 2\lambda m_2(\lambda m_1)^{-1}\|f\|_1 \;=\; 2(m_2/m_1)\|f\|_1,
\end{aligned}
$$

so that condition (2.8) is once again satisfied if $m_2/m_1 < 1/2$.

If, more generally, π_1 is a (signed) compound measure on $\mathbb{Z}$, with characteristic function

$$
\exp\left\{ \lambda \sum_{l\in\mathbb{Z}} \mu_l (e^{il\theta} - 1) \right\},
$$

similar considerations can be applied. Here, we now have $\mathcal{X} = \mathbb{Z}$, but $\mathcal{X}_0$ is still $\mathbb{Z}_+$. The corresponding Stein operator is formally exactly as in (3.5), except that the l-sum now runs over the whole of $\mathbb{Z}$, and we require m_1 to be positive; also, the role of m_2 is now played by $m_2' = \sum_{l\in\mathbb{Z}} l(l-1)|\mu_l|$. When applying Lemma 2.3 and Theorem 2.4, we have the inequalities

$$
\kappa(\pi, \mathbb{Z}_-) \;\le\; 2\,|\pi|(\mathbb{Z}_-)
$$

for use with d_{TV},

$$
\kappa(\pi, \mathbb{Z}_-) \;\le\; \sum_{j<0} |\pi|\{j\}(|j| + \lambda)
$$

for d_W, and, with the fact that $\max_j \pi_0(j) \le (2e\lambda)^{-1/2}$ [Barbour, Holst & Janson (1992, p. 262)],

$$
\kappa(\pi, \mathbb{Z}_-) \;\le\; \frac{1}{\sqrt{2e\lambda}}\,|\pi|(\mathbb{Z}_-) + \max_{l<0} |\pi|\{l\}
$$

for d_{pt}.

Example 3.4. In this example, the setting is similar to that in the preceding example, but we now consider a compound Poisson distribution $\pi_1 = \mathrm{CP}\,(\lambda^1, \mu^1)$ on $\mathbb{Z}_+$ as a perturbation not of a Poisson distribution, but of another compound Poisson distribution $\pi_0 = \mathrm{CP}\,(\lambda^0, \mu^0)$ on $\mathbb{Z}_+$. The reason for doing so is that the solutions to the Stein equation are known to be well behaved only for rather restricted classes of compound Poisson distributions: see Barbour & Utev (1998), Barbour & Xia (2000). The perturbation method offers the possibility of expanding the class of those with good behaviour by including neighbourhoods not only of the Poisson distributions, but also of any other compound Poisson distributions whose Stein solutions can be controlled. In particular, we shall suppose that the distribution $\pi_0 = \mathrm{CP}\,(\lambda^0, \mu^0)$ is such that

$$j\mu_j^0 \;\geq\; (j+1)\mu_{j+1}^0, \quad j \geq 1,$$

and that $\delta := \mu_1^0 - 2\mu_2^0 > 0$, these conditions implying that, with $c_1(\lambda^0) = 4 - 2(\delta\lambda^0)^{-1/2}$ and $c_2(\lambda^0) = \frac{1}{2}(\delta\lambda^0)^{-1} + 2\log^+(2(\delta\lambda^0))$,

$$\|g_f^0\|_\infty \;\leq\; \{\delta\lambda^0\}^{-1/2} c_1(\lambda^0)\|f\|_\infty \quad \text{and} \quad \|\Delta g_f^0\|_\infty \;\leq\; \{\delta\lambda^0\}^{-1} c_2(\lambda^0)\|f\|_\infty, \tag{3.12}$$

where, as usual, $g_f^0 := \mathcal{A}_0^{-1} P_0 f$; see Barbour, Chen & Loh (1992, pp. 1854-5). Here, the Stein operators $\mathcal{A}_0$ and $\mathcal{A}_1$ are given as in (3.5), with the corresponding choices of λ and μ, giving

$$(\mathcal{U}g)(j) \;=\; \sum_{l\geq 1} l\{\lambda^1\mu_l^1 - \lambda^0\mu_l^0\} g(j+l), \quad j \geq 0.$$

As in the previous example, we shall only consider perturbations which preserve the mean, so that also

$$\lambda^1 \sum_{j\geq 1} j\mu_j^1 \;=\; \lambda^0 \sum_{j\geq 1} j\mu_j^0.$$

Taking the supremum norm on $\mathcal{F}$, assumptions (2.1)–(2.4) and (2.6)–(2.7) are satisfied. In order to express the contraction condition (2.8), write

$$E \;:=\; \tfrac{1}{2}\sum_{l\geq 1} l|\lambda^1\mu_l^1 - \lambda^0\mu_l^0|,$$

and define probability measures ρ and σ on $\mathbb{N}$ by

$$\rho_l \;=\; E^{-1} l(\lambda^1\mu_l^1 - \lambda^0\mu_l^0)^+; \quad \sigma_l \;=\; E^{-1} l(\lambda^0\mu_l^0 - \lambda^1\mu_l^1)^+, \qquad l \geq 1;$$

set $\theta := E d_W(\rho, \sigma)$, where d_W denotes the Wasserstein distance. Then, using (3.12), it follows easily that

$$\|\mathcal{U}\mathcal{A}_0^{-1}P_0\|_\infty \;\leq\; \{\delta\lambda^0\}^{-1} c_2(\lambda^0)\theta \;=:\; \gamma,$$

with (2.8) satisfied if $\gamma < 1$.

Example 3.5. In our last example, we consider solving the Stein equation for a point process, whose distribution π_1 is close to that of a spatial Poisson process. Let $\mathbf{X}$ be a compact metric space, and let $\mathcal{X}$ denote the space of Radon measures (point configurations) on $\mathbf{X}$. Then a Poisson process on $\mathbf{X}$ with intensity measure Λ satisfying $\lambda := \Lambda(\mathbf{X}) < \infty$ is a random element $\sum_{l=1}^N \delta_{X_l}$ of $\mathbf{X}$, where $N, X_1, X_2, \ldots$ are all independent, $N \sim \mathrm{Po}\,(\lambda)$ and $X_l \sim \lambda^{-1}\Lambda$ for $l \geq 1$, and δ_x denotes the unit mass at x. Its distribution π_0 can be characterized by the fact that $\pi_0(\mathcal{A}_0 g) = 0$ for all g in

$$\mathcal{G} := \{g\colon \mathcal{X} \to \mathbb{R};\ g(\emptyset) = 0,\ \|\Delta g\|_\infty < \infty\},$$

where

$$(\mathcal{A}_0 g)(\xi) \ := \ \int_{\mathbf{X}} \{(g(\xi + \delta_x) - g(\xi))\Lambda(dx) + (g(\xi - \delta_x) - g(\xi))\xi(dx)\},$$

and $\|\Delta g\|_\infty := \sup_{\xi \in \mathcal{X}, x \in \mathbf{X}} |g(\xi + \delta_x) - g(\xi)|$. Note that the Stein operator $\mathcal{A}_0$ is the generator of a spatial immigration–death process, with π_0 as its equilibrium distribution. For the measure π_1, we take the equilibrium distribution of another spatial immigration–death process on $\mathbf{X}$, with generator $\mathcal{A}_1$ given by

$$(\mathcal{A}_1 g)(\xi) \ := \ \int_{\mathbf{X}} \{(g(\xi + \delta_x) - g(\xi))\Lambda_1(\xi, dx) + (g(\xi - \delta_x) - g(\xi))\xi(dx)\};$$

here, the immigration measure is allowed to depend on the current configuration ξ. We can write $\mathcal{A}_1 = \mathcal{A}_0 + \mathcal{U}$ if we set

$$(\mathcal{U}g)(\xi) \ := \ \int_{\mathbf{X}} (g(\xi + \delta_x) - g(\xi))(\Lambda_1(\xi, dx) - \Lambda(dx)),$$

and we note that $\mathcal{X}_0 := \mathrm{supp}\,(\pi_0) = \mathcal{X}$.

We begin by considering perturbations appropriate for total variation approximation, taking the set of functions $\mathcal{F}\colon \mathcal{X} \to \mathbb{R}$ with the supremum norm $\|\cdot\|_\infty$. Then, as in Barbour & Brown (1992, pp. 12–13), it is possible to define a right inverse $\mathcal{A}_0^{-1}$ satisfying (2.3) and (2.4), with $A = 2$. To check that $\mathcal{U}\mathcal{A}_0^{-1}P_0\colon \overline{\mathcal{F}} \to \overline{\mathcal{F}}$, we combine the definition of $\mathcal{U}$ and (2.4) to give

$$|(\mathcal{U}g_f^0)(\xi)| \ \leq \ 2\overline{L}_\xi(\mathbf{X})\|f\|_\infty,$$

where $\overline{L}_\xi(\cdot)$ denotes the absolute difference between the measures $\Lambda_1(\xi,.)$ and $\Lambda(\cdot)$; hence we shall need in addition to assume that

$$\tilde{\lambda} \;:=\; 2\sup_{\xi\in\mathcal{X}} \overline{L}_\xi(\mathbf{X}) \;<\; \infty, \tag{3.13}$$

in order to make progress. If we do, then (2.8) is satisfied with $\gamma = \tilde{\lambda}$ if $\tilde{\lambda} < 1$, and Theorem 2.4 can be used to show that Stein's method works.

Total variation is often too strong a metric for comparing point process distributions, and so an alternative metric d_2 is proposed in Barbour & Brown (1992), based on test functions Lipschitz with respect to a metric on $\mathbf{X}$ which is bounded by 1. Similar calculations can be carried out in this setting also; the condition needed to satisfy (2.8) is somewhat more stringent.

Even the contraction condition $\tilde{\lambda} < 1$ is rather restrictive. Consider a hard-core model, in which $\mathbf{X} \subset \mathbb{R}^d$ has volume ϑ, $\Lambda(dx) = dx$, and $\Lambda(\xi, dx) = I[\xi(B(x,\varepsilon)) = 0]\,dx$, where $I[C]$ denotes the indicator of the event C; this specification of $\Lambda(\xi, dx)$ is such that no immigration is allowed within distance ε of a point of the current configuration ξ. Then $\tilde{\lambda} = \vartheta$, and contraction is only achieved if the expected number ϑ of points under π_0 is less than 1. However, one could also consider a model with π_1 the equilibrium distribution of a slightly different immigration death process, in which

$$\Lambda(\xi, dx) \;=\; \max\{I[\xi(B(x,\varepsilon)) = 0], I[\xi(\mathbf{X}) > 2\vartheta]\}dx;$$

for large ϑ, the difference between the equilibrium distributions of the two processes is small, but, for the new process, $\tilde{\lambda} \le 2\vartheta a(\varepsilon)$, where $a(\varepsilon)$ is the area of the ε-ball, meaning that models with much larger expected numbers of points can still satisfy the contraction condition. Nonetheless, these are still models in which, at distance ε, little interaction can be expected; the mean number of pairs of points closer than ε to one another under π_0 is about $\frac{1}{2}\vartheta a(\varepsilon)$, and, if the contraction condition is satisfied, this has to be less than $\frac{1}{4}$.

4 Illustrations

In this section, we illustrate how the perturbations described above can be used in specific examples.

Example 4.1. In our first illustration, we return to the setting and notation of Example 3.2, and consider approximation by the distribution π_1 whose Stein operator is given in (3.3) above. The distribution we wish to approximate is the equilibrium distribution π of another jump–diffusion process, in which the jumps do not have fixed size z, but are randomly chosen with z as mean; this process has generator $\mathcal{A}$ given by

$$(\mathcal{A}g)(x) \;=\; g''(x) - xg'(x) + \alpha \int \{g(x+\zeta) - g(x)\}\mu(d\zeta), \quad g \in \mathcal{G}. \quad (4.1)$$

This distribution can be expected to be close to π_1 provided that the probability distribution μ is concentrated about z, and since the distribution π is reasonably well understood, such an approximation may constitute a useful simplification.

The main step is thus to establish a bound of the form (2.18), after which Theorem 2.4 can be applied. However, for $X \sim \pi$ and $g_f^0 := \mathcal{A}_0^{-1}P_0 f$, we immediately have

$$\begin{aligned} \pi(\mathcal{A}_1 g_f^0) \;&=\; \pi(\mathcal{A}g_f^0) - \alpha\mathbb{E}\left\{\int \{g_f^0(X+\zeta) - g_f^0(X+z)\}\mu(d\zeta)\right\} \\ &=\; -\alpha\mathbb{E}\left\{\int \{g_f^0(X+\zeta) - g_f^0(X+z)\}\mu(d\zeta)\right\}, \end{aligned}$$

since $\pi(\mathcal{A}g) = 0$ for all $g \in \mathcal{G}$. From this it follows by the mean value theorem that

$$|\pi(\mathcal{A}_1 g_f^0)| \;\le\; \tfrac{1}{2}\alpha \int (\zeta - z)^2 \mu(d\zeta)\, \|(g_f^0)''\|_\infty \;\le\; 2\alpha \int (\zeta - z)^2 \mu(d\zeta)\, \|f\|_\infty.$$

This suggests that the supremum norm on $\mathcal{F}$ is an appropriate choice, and from Theorem 2.4, if $\gamma := \sqrt{2\pi}\, z\alpha < 1$, as in (3.4), it follows that

$$d_{TV}(\pi, \pi_1) \;\le\; 2\alpha \int (\zeta - z)^2 \mu(d\zeta)/(1-\gamma).$$

Thus the total variation distance between the two distributions is small if the variance of μ is small (and $\gamma < 1$).

Example 4.2. We continue with the setting and notation of Example 3.2, and again approximate by the distribution π_1. Here, as the measure π, we take the equilibrium distribution of a Markov jump process W_N, defined as follows. We let X_N be the pure jump Markov process on $\mathbb{Z}_+$ with transition rates given by

$$\begin{aligned} j &\rightarrow j+1 \quad \text{at rate} \quad N; \qquad j \rightarrow j-1 \quad \text{at rate} \quad j; \\ j &\rightarrow j + \lfloor z\sqrt{N} \rfloor \quad \text{at rate} \quad \alpha, \end{aligned}$$

and we then set $W_N(t) := \{X_N(t) - N\}/\sqrt{N}$. If $z = 0$, the equilibrium distribution of X_N is the Poisson distribution with mean N, and that of W_N the centred and normalized Poisson distribution, which is itself, for large N, close to the normal in Kolmogorov distance, but not in total variation. Here, we wish to find bounds for the accuracy of approximation by π_1 when $z > 0$. As above, we need a bound of the form (2.18), so as to be able to apply Theorem 2.4.

Much as above, we begin by observing that $\pi(\mathcal{A}g) = 0$ for all $g \in \mathcal{G}$, where now, writing $w_{jN} := (j - N)/\sqrt{N}$ and $\eta_N := 1/\sqrt{N}$, we have

$$\begin{aligned} (\mathcal{A}g)(w_{jN}) &= N\{g(w_{jN} + \eta_N) - g(w_{jN})\} \\ &\quad + j\{g(w_{jN} - \eta_N) - g(w_{jN})\} + \alpha\{g(w_{jN} + \lfloor z\sqrt{N}\rfloor \eta_N) - g(w_{jN})\}. \end{aligned}$$

Subtracting $(\mathcal{A}_1 g)(w_{jN})$ and using Taylor's expansion, it follows that

$$\begin{aligned} &|(\mathcal{A}g)(w_{jN}) - (\mathcal{A}_1 g)(w_{jN})| \\ &\qquad \leq N^{-1/2}(\tfrac{1}{3}\|g'''\|_\infty + \tfrac{1}{2}|w_{jN}|\|g''\|_\infty + \alpha\|g'\|_\infty), \end{aligned}$$

so that

$$|\pi(\mathcal{A}_1 g)| \leq N^{-1/2}(\tfrac{1}{3}\|g'''\|_\infty + \tfrac{1}{2}\mathbb{E}|W_N|\|g''\|_\infty + \alpha\|g'\|_\infty). \tag{4.2}$$

Note that, taking $g(w) = w$ and $g(w) = w^2$ respectively in $\pi(\mathcal{A}g) = 0$, as we may, by Hamza & Klebaner (1995, Theorem 2), it follows that $|\mathbb{E}W_N| \leq \alpha z$ and

$$2\mathbb{E}\{W_N^2\} \leq 2N\eta_N^2 + \eta_N|\mathbb{E}W_N| + 2\alpha z|\mathbb{E}W_N| + \alpha z^2 \leq 2 + \alpha z\eta_N + 2\alpha^2 z^2 + \alpha z^2,$$

which implies that

$$\{\mathbb{E}|W_N|\}^2 \leq \mathbb{E}\{W_N^2\} \leq 1 + \tfrac{1}{2}\alpha z\eta_N + \alpha^2 z^2 + \tfrac{1}{2}\alpha z^2;$$

thus $\mathbb{E}|W_N|$ is uniformly bounded in N. Furthermore, for $g = g_f^0 := \mathcal{A}_0^{-1} P_0 f$ and $f \in \overline{\mathcal{F}}^{(1)}$, we can control the first three derivatives of g_f^0 by using Proposition 5.1 with $y = (g_f^0)'$, so that (4.2) yields a bound of the form

$$|\pi(\mathcal{A}_1 g_f^0)| \ \le\ CN^{-1/2} \|f\|^{(1)},$$

for all $f \in \overline{\mathcal{F}}^{(1)}$. In view of Theorem 2.4, this translates into the bound

$$d^{(1)}(\pi, \pi_1) \ \le\ CN^{-1/2}/(1-\gamma)$$

if $\gamma < 1$, where now, for $\|\cdot\|^{(1)}$, we have $\gamma = (4+\sqrt{2\pi})z\alpha$, as in Example 3.2.

If, instead, Kolmogorov distance is of interest, then the only obstacle is to verify (2.25) of Theorem 2.5. For $g = g_h^0$, the estimate given in (4.2) is fine, except for the first term: it is no longer possible to bound the difference

$$D_N(w) \ :=\ N\{g(w+\eta_N) - g(w) + g(w-\eta_N)) - g(w)\} - g''(w)$$

by $\frac{1}{3}\eta_N \|g'''\|_\infty$, since, for $h = h_a = \mathbf{1}_{(-\infty,a]}$, $g'''(a)$ is not defined. However, it is clear that $|D_N(w)| \le 2\|g''\|_\infty$ for all w, and that, for $|w-a| > \eta_N$,

$$|D_N(w)| \ \le\ \sup_{|x-w|\le\eta_N} |g'''(x)|.$$

Now, for $h = h_a$, taking $a > 0$ without real loss of generality, we have

$$|g'''(x)| \ \le\ C_1 + C_2 a e^{-a(a-x)} \mathbf{1}_{(0,a)}(x), \quad x \ne a, \tag{4.3}$$

for universal constants C_1 and C_2, so that g''' is well behaved except just below a. The bound (4.3) can then be combined with the concentration inequality

$$\mathbb{P}[W_N \in [a,b]] \ \le\ \{\tfrac{1}{2}(b-a) + \eta_N\}(\mathbb{E}|W_N| + \alpha z),$$

obtained by taking $g'' = \mathbf{1}_{[a-\eta_N, b+\eta_N]}$ and $g'(w) = \int_{(b-a)/2}^w g''(t)\,dt$ for any $a \le b$ in $\pi(\mathcal{A}g) = 0$, to deduce a bound $\mathbb{E}|D_N(W_N)| \le CN^{-1/2}$, and hence Kolmogorov approximation also at rate $N^{-1/2}$. Total variation approximation is of course never good, since $\mathcal{L}(W_N)$ gives probability 1 to a discrete lattice, and π_1 is absolutely continuous with respect to Lebesgue measure.

Example 4.3. (Borovkov–Pfeifer approximation) Borovkov & Pfeifer (1996) suggested using a single n-independent infinite convolution of simple signed measures as a correction to the Poisson approximation to the distribution of a sum of independent indicator random variables. Their approximation is particularly effective in the case that they treated, the number of records in n i.i.d. trials. Here, the approximation is not as complicated as it might seem, because the generating function of the correcting measure can be conveniently expressed in terms of gamma functions. Its accuracy is then of order $O(n^{-2})$, which is way better than the $O(1/\log n)$ error in the standard Poisson approximation. Their approach was extended to the multivariate case of independent summands in Čekanavičius (2002) and Roos (2003). Note also that Roos (2003) obtained asymptotically sharp constants in the univariate case. In this example, by treating their approximating measure as a perturbation of the Poisson, as in Example 3.3, we investigate Borovkov–Pfeifer approximation to the distribution of the sum of *dependent* Bernoulli random variables.

Let I_i, $i \geq 1$, be dependent Bernoulli $\mathrm{Be}\,(p_i)$ random variables. Define $W = \sum_{i=1}^n I_i$, $W^{(i)} = W - I_i$, and let $\widetilde{W}^{(i)}$ be a random variable having the conditional distribution of $W^{(i)}$ given $I_i = 1$; that is, for all $k \in \mathbb{Z}_+$, $\mathbb{P}(\widetilde{W}^{(i)} = k) = \mathbb{P}(W^{(i)} = k \mid I_i = 1)$. Let

$$\lambda = \sum_{i=1}^n p_i; \qquad \eta_1 = \sum_{i=1}^n \left\{ \frac{p_i}{1-2p_i} \right\} \mathbb{E}|\widetilde{W}^{(i)} - W^{(i)}|.$$

The Borovkov–Pfeifer approximation is defined to be the convolution of the Poisson distribution $\mathrm{Po}\,(\lambda)$ and the signed measure BP determined by its generating function:

$$\widehat{\mathrm{BP}}(z) \;=\; \prod_{i=1}^{\infty} \Big\{ \big(1 + p_i(z-1)\big) \exp\{-p_i(z-1)\} \Big\}. \tag{4.4}$$

Using the fact that

$$\begin{aligned} e^{-p(z-1)}(1+p(z-1)) \;&=\; \exp\{\ln(1+pz/q) - \ln(1+p/q) - p(z-1)\} \\ &=\; \exp\left\{ \frac{p^2}{q}(z-1) + \sum_{l=2}^{\infty} \frac{(-1)^{l+1}}{l} \left(\frac{p}{q}\right)^l (z^l - 1) \right\}, \end{aligned} \tag{4.5}$$

where $q = 1-p$, one can see that BP is a signed compound Poisson measure, provided that $\sum_{i=1}^n p_i^2 < \infty$. Note that $\sum_{i=1}^\infty p_i = \infty$ is allowed, as is indeed the case for record values, when $p_i = 1/i$.

Theorem 4.1 *Assume that $p_i < 1/3$, $i \geq 1$, that $\sum_{i\geq 1} p_i^2 < \infty$, and that*

$$\theta_1 := \frac{m_2'}{m_1} = \frac{\sum_{i=1}^n p_i^2(1-2p_i)^{-2}}{\lambda} < \frac{1}{2}. \tag{4.6}$$

Then

$$d_{TV}(\mathcal{L}(W), \mathrm{Po}(\lambda) * \mathrm{BP}) \leq \frac{2}{\lambda(1-2\theta_1)}\left(\sum_{i=n+1}^{\infty} \frac{p_i^2}{(1-2p_i)^2} + \eta_1\right), \tag{4.7}$$

$$d_{\mathrm{pt}}(\mathcal{L}(W), \mathrm{Po}(\lambda) * \mathrm{BP}) \leq \frac{2}{\lambda(1-2\theta_1)}\left(\sup_k \mathbb{P}(W=k) \sum_{i=n+1}^{\infty} \frac{p_i^2}{(1-2p_i)^2} + \eta_1\right), \tag{4.8}$$

$$d_W(\mathcal{L}(W), \mathrm{Po}(\lambda) * \mathrm{BP}) \leq \frac{1.15}{\sqrt{\lambda}(1-2\theta_1)}\left(\sum_{i=n+1}^{\infty} \frac{p_i^2}{(1-2p_i)^2} + \eta_1\right). \tag{4.9}$$

Remark. Let I_i, $i \geq 1$, be independent. Then it suffices to prove the corresponding approximation for the sum $W_s := \sum_{i=s}^n I_i$ only. Indeed, let BP_s be specified by the generating function:

$$\widehat{\mathrm{BP}_s}(z) = \prod_{i=s}^{\infty}\Big\{\big(1+p_i(z-1)\big)\exp\{-p_i(z-1)\}\Big\}.$$

Then

$$\mathrm{Po}(\lambda) * \mathrm{BP} = \mathcal{L}\left(\sum_{i=1}^{s-1} I_i\right) * \mathrm{Po}\left(\sum_{i=s}^{n} p_i\right) * \mathrm{BP}_s$$

and

$$\mathcal{L}(W) = \mathcal{L}\left(\sum_{i=1}^{s-1} I_i\right) * \mathcal{L}(W_s),$$

and so, by the properties of total variation we have

$$d_{TV}\left(\mathcal{L}(W), \mathrm{Po}(\lambda) * \mathrm{BP}\right) \leq d_{TV}\left(\mathcal{L}(W_s), \mathrm{Po}\left(\sum_{i=s}^{n} p_i\right) * \mathrm{BP}_s\right).$$

If W is the sum of independent Bernoulli variables, then $\eta_1 = 0$ and

$$\sup_k \mathbb{P}(W=k) \leq \left(4\sum_{i=1}^{n} p_i(1-p_i)\right)^{-1/2},$$

see Barbour & Jensen (1989, Lemma 1). Now, if we consider the records example of Borovkov & Pfeifer (1996), with $p_i = 1/i$, we can take any $s \geq 4$ in the remark above, and obtain orders of accuracy for the total variation distance, point metric and Wasserstein metric of $O((n \ln n)^{-1})$, $O(n^{-1}(\ln n)^{-3/2})$ and $O(n^{-1}(\ln n)^{-1/2})$, respectively.

Proof of Theorem 4.1. In this case, $\mathcal{X} = \mathcal{X}_0 = \mathbb{Z}_+$. Using (4.5), and setting $q_i = 1 - p_i$, we can write $\mathrm{Po}\,(\lambda) * \mathrm{BP}$ as the signed compound Poisson measure with generating function

$$\exp\left\{\sum_{l\geq 1} \lambda_l (z^l - 1)\right\}, \tag{4.10}$$

where $\lambda_l = \lambda_{1l} + \lambda_{2l}$, with

$$\begin{aligned} \lambda_{1l} &= \frac{(-1)^{l+1}}{l} \sum_{i=1}^{n} \left(\frac{p_i}{q_i}\right)^l, \quad l \geq 1; \\ \lambda_{21} &= \sum_{i=n+1}^{\infty} \frac{p_i^2}{q_i}; \qquad \lambda_{2l} = \frac{(-1)^{l+1}}{l} \sum_{i=n+1}^{\infty} \left(\frac{p_i}{q_i}\right)^l, \quad l \geq 2. \end{aligned}$$

Here, the components λ_{1l} come from the signed compound Poisson representation of a sum of independent Bernoulli $\mathrm{Be}\,(p_i)$ random variables, $1 \leq i \leq n$, and the λ_{2l} from the remaining BP_{n+1} measure.

Let $\mu_l = \lambda_l/\lambda$. Then, since $\sum_{l=1}^{\infty} l\lambda_{1l} = \sum_{i=1}^{n} p_i = \lambda$ and $\sum_{l=1}^{\infty} l\lambda_{2l} = 0$, we have $m_1 = \sum_{l=1}^{\infty} l\mu_l = 1$. Hence, the formula for θ_1 follows directly from

$$\sum_{l=2}^{\infty} l(l-1)|\lambda_l| = \sum_{l=2}^{\infty} (l-1) \sum_{i=1}^{\infty} \left(\frac{p_i}{q_i}\right)^l = \sum_{i=1}^{\infty} p_i^2 (1-2p_i)^{-2}.$$

Next, we take Stein operators $\mathcal{A}_0$ as in (3.6) and $\mathcal{A}_1$ as in (3.5). For $g = g_f^0 := \mathcal{A}_0^{-1} P_0 f$, it follows that

$$\mathbb{E}(\mathcal{A}_1 g)(W) = \left\{\sum_{l=1}^{\infty} l\lambda_{1l}\,\mathbb{E}\,g(W+l) - \mathbb{E}\{W g(W)\}\right\} + \sum_{l=1}^{\infty} l\lambda_{2l}\,\mathbb{E}\,g(W+l). \tag{4.11}$$

We begin by bounding the quantity in braces, which gives a bound for the accuracy of the approximation of $\mathcal{L}(W)$ by the distribution of a sum of independent Bernoulli Be$\,(p_i)$ random variables. We observe immediately that, for any i and l,

$$\mathbb{E}g(W+l) \;=\; p_i\mathbb{E}g(\widetilde{W}^{(i)}+l+1) + q_i\mathbb{E}\{g(W^{(i)}+l)\,|\,I_i=0\}$$

and that

$$\mathbb{E}g(W^{(i)}+l) \;=\; p_i\mathbb{E}g(\widetilde{W}^{(i)}+l) + q_i\mathbb{E}\{g(W^{(i)}+l)\,|\,I_i=0\},$$

from which it follows that

$$\mathbb{E}g(W+l) \;=\; q_i\mathbb{E}g(W^{(i)}+l) + p_i\mathbb{E}g(\widetilde{W}^{(i)}+l+1) + p_i u_{il},$$

where we write $u_{il} := \mathbb{E}g(W^{(i)}+l) - \mathbb{E}g(\widetilde{W}^{(i)}+l)$. Setting $v_{il} := (-1)^{l+1}(p_i/q_i)^l$, so that $l\lambda_{1l} = \sum_{i=1}^n v_{il}$, and observing that $p_i v_{il} = -q_i v_{i,l+1}$, we thus have

$$\begin{aligned}
&\sum_{l\ge1} v_{il}\mathbb{E}g(W+l) - \mathbb{E}\{I_i g(W)\}\\
&\quad= \; q_i\sum_{l\ge1} v_{il}\mathbb{E}g(W^{(i)}+l) - q_i\sum_{l\ge2} v_{il}\mathbb{E}g(\widetilde{W}^{(i)}+l)\\
&\qquad\quad + p_i\sum_{l\ge1} v_{il}u_{il} - p_i\mathbb{E}g(\widetilde{W}^{(i)}+1)\\
&\quad= \; \sum_{l\ge1} v_{il}u_{il}.
\end{aligned}$$

Adding over $1 \le i \le n$, we thus find that

$$\begin{aligned}
&\left|\sum_{l=1}^{\infty} l\lambda_{1l}\,\mathbb{E}g(W+l) - \mathbb{E}\{Wg(W)\}\right|\\
&\qquad\le\; \sum_{i=1}^n\sum_{l\ge1} |v_{il}|\,|\mathbb{E}g(W^{(i)}+l) - \mathbb{E}g(\widetilde{W}^{(i)}+l)| \;\le\; \eta_1\|\Delta g\|_\infty.
\end{aligned}$$

It now remains to estimate the remaining element $\sum_{l=1}^{\infty} l\lambda_{2l}\,\mathbb{E}g(W+l)$ in (4.11). Using the identity

$$g(W+l) \;=\; g(W+1) + \sum_{s=1}^{l-1}\Delta g(W+s),$$

we have

$$
\begin{aligned}
\sum_{l=1}^{\infty} l\lambda_{2l}\,\mathbb{E}g(W+l) &= \mathbb{E}g(W+1)\left\{\sum_{l=1}^{\infty} l\lambda_{2l}\right\} + \sum_{l=1}^{\infty} l\lambda_{2l}\sum_{s=1}^{l-1}\mathbb{E}\{\Delta g(W+s)\} \\
&= \sum_{l=1}^{\infty} l\lambda_{2l}\sum_{s=1}^{l-1}\mathbb{E}\{\Delta g(W+s)\},
\end{aligned}
$$

because $\sum_{l=1}^{\infty} l\lambda_{2l} = 0$. Now we have

$$
\begin{aligned}
|\mathbb{E}\{\Delta g(W+s)\}| &\le \min\left\{\|\Delta g\|_\infty,\ \|\Delta g\|_1 \max_k \mathbb{P}(W=k)\right\}, \\
\sum_{l=2}^{\infty} l(l-1)|\lambda_{2l}| &= \sum_{i=n+1}^{\infty} \frac{p_i^2}{(1-2p_i)^2}.
\end{aligned}
$$

The estimates (4.7)–(4.9) thus follow directly from the inequalities (3.8), (3.10) and (3.11) in Example 3.3. □

5 Appendix

Here, we collect various properties of the solution y to the equation

$$
y'(x) - (1-\psi)xy(x) = h(x) - \bar{h}_\psi, \quad x \in \mathbb{R}, \tag{5.1}
$$

for given h and $0 \le \psi < 1$, where $\bar{h}_\psi = \mathbb{E}h(N)$, for $N \sim \mathcal{N}(0,(1-\psi)^{-1})$.

Proposition 5.1

(a) *If $h = \mathbf{1}_{(-\infty,z]}$ for any $z \in \mathbb{R}$, then*

$$
\begin{aligned}
&\text{(i)} \quad \|y\|_\infty \le \frac{1}{4}\sqrt{\frac{2\pi}{1-\psi}}; \\
&\text{(ii)} \quad \|y'\|_\infty \le 1; \\
&\text{(iii)} \quad \sup_x |xy(x)| \le \frac{1}{1-\psi}.
\end{aligned}
$$

(b) *If h is bounded, then*

$$\begin{aligned}
&\text{(i)} \qquad \|y\|_\infty \;\le\; \sqrt{\frac{2\pi}{1-\psi}}\,\|h\|_\infty; \\
&\text{(ii)} \qquad \|y'\|_\infty \;\le\; 4\,\|h\|_\infty; \\
&\text{(iii)} \qquad \sup_x |xy(x)| \;\le\; \frac{2}{1-\psi}\,\|h\|_\infty.
\end{aligned}$$

(c) *If h is uniformly Lipschitz, then*

$$\begin{aligned}
&\text{(i)} \qquad \|y\|_\infty \;\le\; \frac{2}{1-\psi}\,\|h'\|_\infty; \\
&\text{(ii)} \qquad \|y'\|_\infty \;\le\; \frac{4}{\sqrt{1-\psi}}\,\|h'\|_\infty; \\
&\text{(iii)} \qquad \|y''\|_\infty \;\le\; \frac{2}{\sqrt{1-\psi}}\,\|h'\|_\infty; \\
&\text{(iv)} \qquad \sup_x |xy'(x)| \;\le\; \frac{3}{1-\psi}\,\|h'\|_\infty.
\end{aligned}$$

Proof. Equation (5.1) can be transformed, using the substitution $x = w/\sqrt{1-\psi}$, into the equation with $\psi = 0$ for the standard normal distribution, for which the corresponding bounds are mostly given in Chen & Shao (2005, Lemmas 2.2 and 2.3). In particular, the bounds (a)(i)–(iii) follow directly from their Equations (2.9), (2.8) and (2.7), respectively; the bounds (b)(i)–(ii) from the proofs of their Equations (2.11) and (2.12); and the bounds (c)(i)–(iii) from their Equations (2.11)–(2.13).

The bound (b)(iii) is easily deduced from the explicit expression for the solution y: for instance, for $x > 0$, we have

$$xy(x) \;=\; -xe^{(1-\psi)x^2/2}\int_x^\infty e^{-(1-\psi)t^2/2}(h(t) - \bar{h}_\psi)\,dt\,,$$

immediately giving

$$|xy(x)| \;\le\; x\int_0^\infty e^{-(1-\psi)zx}\,|h(x+z) - \bar{h}_\psi|\,dz\,,$$

from which (b)(iii) follows.

For (c)(iv), we argue only for $x < 0$, since the proof for $x > 0$ is entirely similar. Noting that

$$y''(x) - (1-\psi)xy'(x) \;=\; (1-\psi)y(x) + h'(x)\,,$$

we obtain

$$y'(x) \;=\; e^{\frac{(1-\psi)x^2}{2}} \int_{-\infty}^{x} \{(1-\psi)y(t) + h'(t)\}\, e^{-\frac{(1-\psi)t^2}{2}}\, dt\,;$$

hence

$$\begin{aligned} |xy'(x)| \;&\le\; \{(1-\psi)\|y\|_\infty + \|h'\|_\infty\}|x|e^{\frac{(1-\psi)x^2}{2}} \int_{-\infty}^{x} e^{-\frac{(1-\psi)t^2}{2}}\, dt \\ &\le\; \|y\|_\infty + \|h'\|_\infty/(1-\psi)\,. \end{aligned}$$

But now, from (c)(i) above, $\|y\|_\infty \le \frac{2}{1-\psi}\,\|h'\|_\infty\,.$ $\square$

Acknowledgement. We wish to thank a referee for extremely helpful suggestions as to presentation. We also wish to acknowledge the generous support that has made this research possible; in particular from Schweizerischer Nationalfonds Projekt Nr. 20-107935/1 and from the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems (ADB) and from ARC Discovery Grant DP0209179 (VC).

References

[1] Barbour, A. D. & Brown, T. C. (1992). Stein's method and point process approximation, *Stochastic Processes and their Applications* **43** 9-31.

[2] Barbour, A. D. & Čekanavičius, V. (2002). Total variation asymptotics for sums of independent integer random variables, *Ann. Probab.* **30** 509-545.

[3] Barbour, A. D., Chen, L. H. Y. & Loh, W. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method, *Ann. Probab.* **20** 1843-1866.

[4] Barbour, A. D., Holst, L. & Janson, S. (1992). *Poisson Approximation,* Oxford Univ. Press.

[5] Barbour, A. D. & Jensen, J. L. (1989) Local and tail approximations near the Poisson limit. *Scandinavian J. Statist.* **16** 75-87.

[6] Barbour, A. D. & Utev, S. (1998). Solving the Stein equation in compound Poisson approximation. *Adv. Appl. Probab.* **30** 449-475.

[7] Barbour, A. D. & Xia, A. (1999). Poisson Perturbations. *ESAIM: Probab. Stat.* **3** 131-150.

[8] Barbour, A. D. & Xia, A. (2000). Estimating Stein's constants for compound Poisson approximation. *Bernoulli* **6** 581–590.

[9] Barbour, A. D. & Xia, A. (2005). On Stein's factors for Poisson approximation in Wasserstein distance. Preprint.

[10] Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

[11] Borovkov, K. A. & Pfeifer, D. (1996). On improvements of the order of approximation in the Poisson limit theorem. *J. Appl. Prob.* **33** 146-155.

[12] Cacoullos, T., Papathanasiou, V. & Utev, S. A. (1994) Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22** 1607–1618.

[13] Čekanavičius, V. (2002) On multivariate compound distributions. *Teorya Veroyatn. Primen.* **47** 583-594.

[14] Čekanavičius, V. (2004). On local estimates and the Stein method. *Bernoulli.* **10** 665-683.

[15] Chen, L. H. Y. & Shao, Q. M. (2005). Stein's method for normal approximation. In: *An Introduction to Stein's Method,* Eds. A. D. Barbour & L. H. Y. Chen, World Scientific Press, Singapore.

[16] Hamza, K. & Klebaner, F. C. (1995). Conditions for integrability of Markov chains. *J. Appl. Probab.* **32** 541-547.

[17] Roos, B. (2003). Poisson approximation via the convolution with Kornya-Presman signed measures. *Theor. Probab. Appl.* **48**(3) 555-560.

[18] Stein, C. (1972). A bound error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkley Symp. Math. Statist. Probab.* **3** 583-602. Univ. California Press, Berkley.